

Metabolic signature of healthy lifestyle and its relation with risk of hepatocellular carcinoma in a large European cohort

Nada Assi,¹ Marc J Gunter,¹ Duncan C Thomas,² Michael Leitzmann,³ Magdalena Stepien,¹ Véronique Chajès,¹ Thierry Philip,⁴ Paolo Vineis,⁵ Christina Bamia,^{6,7} Marie-Christine Boutron-Ruault,^{8,9} Torkjel M Sandanger,¹⁰ Amaia Molinuevo,^{11,12} Hendriek Boshuizen,¹³ Anneli Sundkvist,¹⁴ Tilman Kühn,¹⁵ Ruth Travis,¹⁶ Kim Overvad,¹⁷ Elio Riboli,⁵ Augustin Scalbert,¹ Mazda Jenab,¹ Vivian Viallon,^{1,18} and Pietro Ferrari¹

¹Section of Nutrition and Metabolism, International Agency for Research on Cancer (IARC), Lyon, France; ²University of Southern California, Los Angeles, CA; ³Department of Epidemiology and Preventive Medicine, Regensburg University, Regensburg, Germany; ⁴Unité Cancer et Environnement, Centre Léon Bérard, Lyon, France; ⁵Department of Epidemiology and Biostatistics, MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom; ⁶Hellenic Health Foundation, Athens, Greece; ⁷WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece; ⁸Université Paris-Saclay, Université Paris-Sud, Villejuif, France; ⁹Gustave Roussy, Villejuif, France; ¹⁰Department of Community Medicine, UiT the Arctic University of Norway, Tromsø, Norway; ¹¹Public Health Division of Gipuzkoa, Regional Government of the Basque Country, Donostia-San Sebastián, Spain; ¹²CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain; ¹³National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, Netherlands; ¹⁴Department of Radiation Sciences Oncology, Umeå University 901 87 Umeå, Sweden; ¹⁵Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; ¹⁶Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom; ¹⁷The Department of Epidemiology, School of Public Health, Aarhus University, Aarhus, Denmark; and ¹⁸Université de Lyon, Université Claude Bernard Lyon1, Lyon, France

ABSTRACT

Background: Studies using metabolomic data have identified metabolites from several compound classes that are associated with disease-related lifestyle factors.

Objective: In this study, we identified metabolic signatures reflecting lifestyle patterns and related them to the risk of hepatocellular carcinoma (HCC) in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort.

Design: Within a nested case-control study of 147 incident HCC cases and 147 matched controls, partial least squares (PLS) analysis related 7 modified healthy lifestyle index (HLI) variables (diet, BMI, physical activity, lifetime alcohol, smoking, diabetes, and hepatitis) to 132 targeted serum-measured metabolites and a liver function score. The association between the resulting PLS scores and HCC risk was examined in multivariable conditional logistic regression models, where ORs and 95% CIs were computed.

Results: The lifestyle component's PLS score was negatively associated with lifetime alcohol, BMI, smoking, and diabetes, and positively associated with physical activity. Its metabolic counterpart was positively related to the metabolites sphingomyelin (SM)(OH) C14:1, C16:1, and C22:2 and negatively related to glutamate, hexoses, and the diacyl-phosphatidylcholine PC aaC32:1. The lifestyle and metabolomics components were inversely associated with HCC risk, with the ORs for a 1-SD increase in scores equal to 0.53 (95% CI: 0.38, 0.74) and 0.28 (0.18, 0.43), and the associated AUCs equal to 0.64 (0.57, 0.70) and 0.74 (0.69, 0.80), respectively.

Conclusions: This study identified a metabolic signature reflecting a healthy lifestyle pattern which was inversely associated with HCC risk. The metabolic profile displayed a stronger association

with HCC than did the modified HLI derived from questionnaire data. Measuring a specific panel of metabolites may identify strata of the population at higher risk for HCC and can add substantial discrimination compared with questionnaire data. This trial was registered at clinicaltrials.gov as NCT03356535. *Am J Clin Nutr* 2018;108:117–126.

Keywords: hepatocellular carcinoma, targeted metabolomics, multivariate statistics, metabolic signatures, partial least squares, healthy lifestyle index, EPIC

INTRODUCTION

Hepatocellular carcinoma (HCC) is the predominant form of liver cancer and is the second most frequent cause of cancer death worldwide (1). HCC incidence rates have risen dramatically in Europe in the recent decades (2), along with HCC mortality, with unfavorable trends projected to 2020 (3). HCC is a multifactorial disease strongly associated with lifestyle factors (4). Although hepatitis infection remains its primary risk factor, other exposures, such as obesity, alcohol drinking, diabetes, smoking, physical activity, and some dietary factors, have been related to HCC risk (5–7). The rising prevalence of obesity, diabetes, and alcohol drinking may explain the observed trends in HCC incidence, particularly in regions where hepatitis infection rates are low (4, 8). The prevalence of hepatitis B virus is estimated to be ~0.9% and that of hepatitis C virus ~1.1% in the European Union (compared with 1% and 0.8%, respectively, in the United States) (9, 10).

The etiology of HCC is complex, and the pathways implicated in hepatocarcinogenesis entail a broad range of mechanisms that are likely affected by more than one exposure (2). Better understanding of this process calls for a comprehensive approach to exploring HCC risk factors, in particular investigating the effects of multiple modifiable lifestyle factors as well as their biological correlates. Studies investigating etiologic determinants of HCC often rely on variables ascertained in self-administered questionnaires that may be subject to various biases, including measurement error and recall bias (11). The identification and application of biomarkers in this regard might not only improve measurement of the relevant etiological determinants, but may also help elucidate the link between lifestyle exposures and disease progression by revealing insights into the underlying biological mechanisms (12, 13).

Metabolomics is an evolving methodology that measures a broad spectrum of small molecules to identify causal and mechanistic pathways in disease development and etiology (14). Recently, epidemiologic studies incorporating metabolomic data have identified metabolites related to a number of disease outcomes, including different cancer subtypes (15, 16), diabetes (17), alcoholic hepatitis (18), and hepatobiliary disease (12). A number of investigations have also found metabolic

The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (Netherlands); Nordic Centre of Excellence Programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS), PI13/00061 to Granada, Regional Governments of Andalucía, Asturias, Basque Country, Murcia (no. 6236) and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Scientific Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk and MR/M012190/1 to EPIC-Oxford) (United Kingdom).

This study was done independently and with no input from the funders. The funders were not involved in the design, implementation, analysis, or interpretation of the data.

Supplemental Methods, Supplemental Table 1 and 2, and Supplemental Figures 1–3 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/ajcn/>.

Address correspondence to PF (e-mail: ferrari@iarc.fr).

Abbreviations used: EM, expectation maximization; EPIC, European Prospective Investigation into Cancer and Nutrition; HCC, hepatocellular carcinoma; HLI, healthy lifestyle index; IARC, International Agency for Research on Cancer; LysoPC, lyso-phosphatidylcholine; PCA, principal component analysis; PC aa, diacyl-phosphatidylcholine; PC ae, acyl-alkyl-phosphatidylcholine; PC-PR2, Principal Component Partial R2; PLS, partial least squares; ROC, receiver operating characteristic curve; SM, sphingomyelin; T2D, type 2 diabetes.

Received November 29, 2017. Accepted for publication March 21, 2018.

First published online June 18, 2018; doi: <https://doi.org/10.1093/ajcn/nqy074>.

markers associated with dietary (19, 20) and lifestyle exposures, including physical activity (19), obesity (21), alcohol consumption (13, 22), and smoking (23). These exposures are established HCC risk factors and are components of an established healthy lifestyle index (HLI), the adherence to which has been associated with lower risks of noncommunicable diseases, including cancer (24, 25).

In this study, the lifestyle variables of a modified HLI (24, 25), including diet, BMI, physical activity, smoking, alcohol intake, diabetes, and hepatitis infection, were related to specific metabolic signals acquired through targeted metabolomics. Metabolic signatures reflecting healthy lifestyle behaviors were identified and related to HCC risk within a nested case-control study of HCC in the European Prospective Investigation into Cancer and Nutrition (EPIC).

METHODS

Study population

The EPIC is a multinational prospective cohort study designed to investigate the link between diet, lifestyle, and environmental factors with cancer incidence and other chronic disease outcomes. Over 520,000 healthy men and women aged 25–85 y were enrolled between 1992 and 2000 across 23 EPIC administrative centers in 10 European countries: Denmark, France, Germany, Greece, Italy, Netherlands, Norway, Spain, Sweden, and the United Kingdom (26). In most of the EPIC centers, participants were recruited from the general population with the following exceptions: for France, women were enrolled from a health insurance scheme for school and university employees; in Utrecht (Netherlands) and in Florence (Italy), participants came from breast cancer screening programs; some centers in Italy (Turin and Ragusa) and Spain recruited blood donors; and the Oxford subcohort (United Kingdom) included mostly health-conscious individuals recruited throughout the United Kingdom. Finally, the French, Norwegian, and Naples (Italy) cohorts comprised only women. Extensive details of the study design and recruitment methods have been published previously (26).

Collection of dietary and lifestyle data

During the enrollment period, participants gave informed consent and completed questionnaires on diet, lifestyle, and medical history. Biological samples were collected for approximately 80% of the cohort prior to disease onset. Serum samples were stored at the International Agency for Research on Cancer (IARC), Lyon, France in -196°C liquid nitrogen for all countries, with the exception of the samples originating from Sweden (-80°C freezers) and Denmark (-150°C nitrogen vapor). Usual diet over the previous 12 mo was assessed for each individual through validated country-specific dietary questionnaires (26). Nutrient intakes were then estimated with the use of a common harmonized food composition database across EPIC countries (EPIC Nutrient Database, ENDB) (27). Information on sociodemographic data, including education, smoking, and alcohol drinking histories, as well as physical activity were gathered in lifestyle questionnaires. Anthropometric characteristics were directly measured by trained study personnel for most of the participants (26), but were self-reported in baseline questionnaires for a subset of participants from the EPIC-Oxford subcohort,

although the accuracy of these self-reported data have been validated (28). Approval for this study was obtained from the ethical review boards of the participating institutions and IARC.

Nested case-control study of HCC

Within a nested case-control study of HCC (29) in EPIC, this analysis focused on 147 cases and 147 matched controls with available biological samples identified in the period between the subjects' recruitment into the cohort and 2010 (29) (**Supplemental Figure 1**). For each HCC case, one control ($n = 147$) was selected by incidence density sampling from all cohort members alive and free of cancer (except for nonmelanoma skin cancer), and matched by age at blood collection (± 1 y), sex, study center, time of the day at blood collection (± 3 h), and fasting status at blood collection (<3 , 3 – 6 , and >6 h); among women, the pair was additionally matched by menopausal status (pre-, peri-, and postmenopausal) and hormone replacement therapy use at time of blood collection (yes/no). Cases of HCC originated from all participating EPIC centers except for Norway and France, which were not included in this study. All subjects were cancer free at the time of blood collection.

Follow-up and case ascertainment in the nested case-control study

Follow-up started at date of entry to the study and finished at date of diagnosis, death, or last completed follow-up (from December 2004 up to June 2010), whichever came first. Cancer incidence was determined through population cancer registries or through active follow-up, as detailed elsewhere (29). Incident HCC cases were defined as first primary invasive tumours and identified through the 10th Revision of International Statistical Classification of Diseases, Injury and Causes of Death as C22.0, with morphology codes ICD-O-2 "8170/3" and "8180/3." Metastatic cases and other types of primary liver cancer were excluded.

Lifestyle variables

The panel of lifestyle variables included in this analysis were BMI (continuous, kg/m^2), average lifetime alcohol intake (continuous, grams per day), the diet score (continuous) described in the **Supplemental Methods**, physical activity (continuous metabolic equivalents of task in metabolic equivalent hours per week), smoking (never, former smokers quit >10 y, former smokers quit ≤ 10 y, current smokers ≤ 15 cigarettes/d, current smokers >15 cigarettes/d), hepatitis infection (yes/no), and self-reported diabetes at baseline (yes/no). These are the components of an HLI used in EPIC (24, 25), modified to include hepatitis and diabetes status, as detailed in Supplemental Methods and **Supplemental Table 1**. Average lifetime alcohol intake was used instead of alcohol intake at recruitment to address potential bias related to reverse causality.

These seven lifestyle variables are referred to herein as the X-set. Missing values in the X-set were imputed through a simple expectation maximization (EM) algorithm using the covariance matrix of the data (30).

Metabolomic data

Concentrations of a predefined set of common endogenous metabolic biomarkers were measured in serum samples at IARC, Lyon, France, with the use of a BIOCRATES AbsoluteIDQ p180 Kit (Biocrates), with ultra-performance liquid chromatography (1290 Series HPLC, Agilent) coupled to a tandem mass spectrometer (QTrap 5500, AB Sciex). Details of the sample preparation methods and mass spectrometry analyses are provided elsewhere (29). Platform reliability has been previously assessed with ICCs above 0.50 in 73% and 52% of the metabolites in fasting and nonfasting samples, respectively (31). Metabolites with CVs $>20\%$ for analytical replicates were excluded, resulting in 145 detected metabolites. Of these, metabolites with $>40\%$ of missing values were excluded (i.e., values that are below the limit of detection/quantification or above the highest calibration standards), resulting in a total of 132 metabolites being included in the study. Measurements that were below the limit of detection were set to half that value and those below the limit of quantification were set to half that limit (applicable to a total of 16 metabolites for 0.3–29.3% of participants). Additionally, measurements that were above the highest concentration calibration standards were set to the highest values. The metabolite nomenclature has been described previously (21), and can be found in Supplemental Methods.

Liver function score

A composite score indicative of liver function identifying the number of abnormal values for six circulating liver blood biomarker tests indicating possible underlying liver dysfunction (32) was included in the set of metabolites. The score entailed the following tests: alanine aminotransferase >55 U/L, aspartate aminotransferase >34 U/L, gamma-glutamyltransferase: men >64 U/L and women >36 U/L, alkaline phosphatase >150 U/L, albumin <35 g/L, and total bilirubin >20.5 $\mu\text{mol}/\text{L}$. The cut-off points were provided by the clinical biochemistry laboratory that conducted the analyses (Centre de Biologie République, Lyon, France) based on assay specifications, as previously described (32). These biomarkers were acquired at the same time as the metabolites from the prediagnostic blood samples collected at recruitment.

The set comprising the 132 metabolites and the liver function score is referred to as the M-set, containing the metabolomic data.

Statistical analyses

Identification of sources of systematic variability within the data

To identify and quantify the sources of systematic variability within the X-set of HLI variables and the M-set of metabolites, the Principal Component Partial R2 (PC-PR2) method was used (33). In this study, PC-PR2 was applied to the X-set of 7 exposure variables, with the covariates explored for systematic variability being country, age at recruitment, and sex. With the similar objective of identifying sources of variability in the metabolomic data, another PC-PR2 analysis was run on the M-set examining the contribution to variability of the following covariates: country, age at blood collection, batch, sex, BMI, diet score, physical activity, alcohol at recruitment, smoking, hepatitis, and diabetes at baseline. PC-PR2 combines aspects of principal

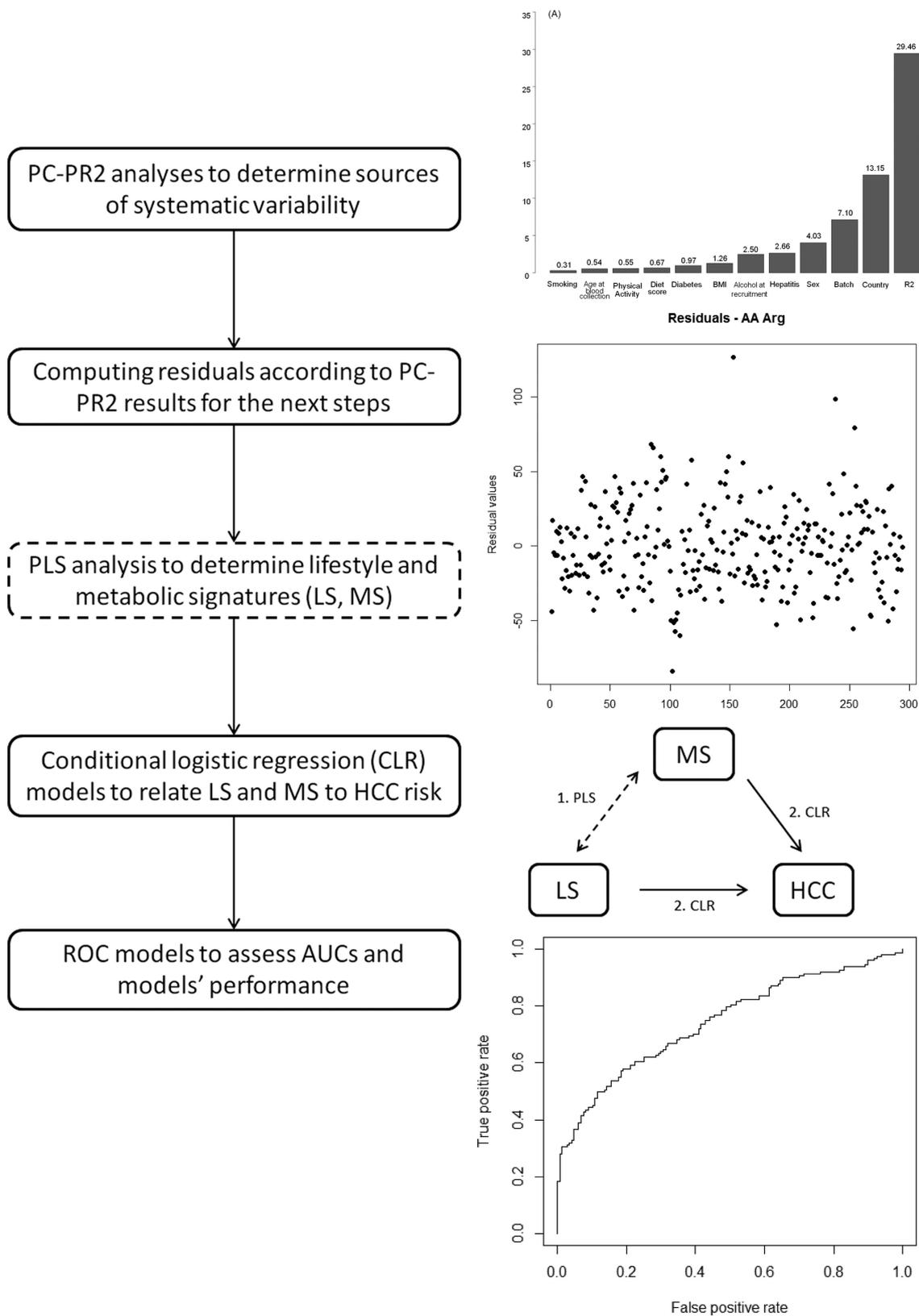


FIGURE 1 The different analytical steps conducted in this study. CLR, conditional logistic regression; HCC, hepatocellular carcinoma; LS, lifestyle signatures; MS, metabolic signatures; PC-PR2, Principal Component Partial R2; PLS, partial least square; ROC, receiver operating characteristic curve.

TABLE 1Baseline characteristics of the study population of the EPIC nested case-control study on hepatocellular carcinoma¹

Characteristics	Cases	Controls
<i>n</i>	147	147
Sex, <i>n</i>		
Male	102	102
Female	45	45
Lag time to cancer diagnosis, y	6.35 (1.18, 10.26)	—
Age at blood collection, y	60.1 (50.7, 68.8)	60.1 (50.9, 68.9)
Height, cm	167.7 (152.3, 180.7)	169.3 (156.1, 181.0)
Weight, kg	79.8 (59.0, 102.2)	78.3 (60.6, 93.0)
BMI, kg/m ²	28.2 (23.0, 34.9)	27.3 (22.0, 32.5)
Total energy, kcal/d	2260.8 (1381.4, 3169.3)	2276.6 (1495.1, 3140.6)
Country		
Denmark	25	25
Germany	32	32
Greece	16	16
Italy	28	28
Netherlands	4	4
Spain	11	11
Sweden	16	16
United Kingdom	15	15
Alcohol intake at recruitment, g/d		
Among consumers	29.6 (1.08, 80.76)	16.8 (1.27, 41.27)
% of never drinkers (<0.1 g/d)	13.6	6.1
Physical activity, metabolic equivalent h/wk	77.9 (18.9, 150.5)	83.3 (21.6, 157.2)
Lifetime alcohol consumption, ² g/d		
Among consumers	33.6 (2.13, 74.42)	19.7 (2.17, 41.32)
% of never drinkers (<0.1 g/d)	3.1	4.1
Diet score ²	25.7 (16.0, 33.0)	27.4 (20.6, 30.0)
Hepatitis infection ^{2,3}		
Yes	41	5
No	106	142
Diabetes at baseline ²		
Yes	19	10
No	128	137
Smoking status ²		
Current smokers, >15 cigarettes/d	25	23
Current smokers, ≤15 cigarettes/d	34	10
Former smokers, quit ≤10 y	17	25
Former smokers, quit >10y	29	29
Never	42	60

¹ Values are presented as means with 10th and 90th percentiles in parentheses for continuous variables and as frequencies for categorical variables. EM, expectation maximization; EPIC, European Prospective Investigation into Cancer and Nutrition.

² There were respectively 42, 12, 76, 29, and 7 missing values for lifetime alcohol consumption, diet score, hepatitis, diabetes, and smoking; they were imputed with an EM algorithm using the covariance matrix of the data.

³ Prior to EM, there were 41 hepatitis infections (3 in controls and 38 in cases). After imputation, there were 46 hepatitis infections (5 in controls and 41 in cases). Of the initial distribution of the 41 hepatitis patients, 26 had hepatitis C and 15 had hepatitis B, with 3 having both hepatitis B virus and hepatitis C virus.

component analysis (PCA) with the partial R^2 statistic in multiple linear regression models. Briefly, PCA is performed on the set under scrutiny, and a number of components explaining an amount of total variability above a designated threshold (here, 80%) are retained. Multiple linear models are then fitted, where each component's variability is explained by regressing it on a list of relevant covariates, yielding an R^2 statistic for each of the latter. The R^2 quantifies the amount of variability each independent variable explained, conditionally on all other covariates included in the model. Finally, an overall partial R^2 is computed as a weighted mean for each covariate, with the eigenvalues as the components' weights. Subsequently, residuals on the variables accounting for

most variability in lifestyle and metabolomic data were computed in a series of univariate linear regression models and were used in the following partial least squares (PLS) analyses.

Lifestyle pattern and metabolic signature assessment

With the aim of deriving lifestyle and metabolic signatures that mirror one another while relating the entire set of exposure variables to the metabolomic data, we used PLS analysis, a multivariate method that achieves dimensionality reduction. PLS generalizes features of PCA with those of multiple linear

TABLE 2
Exposure lifestyle variables and corresponding metabolites contributing to the first PLS factor¹

Exposure variable	Loadings	Metabolites	Loadings ²
Main PLS analysis³			
BMI	−0.385	Glutamic acid	−0.192
Lifetime alcohol	−0.695	Hexoses	−0.191
Diet score	−0.058	SM(OH) C14:1	0.196
Physical activity	0.297	SM(OH) C16:1	0.179
Smoking	−0.409	SM(OH) C22:2	0.214
Hepatitis infection	−0.176	PC aaC32:1	−0.184
Diabetes	−0.282	Liver function score	−0.186
Sensitivity PLS analysis⁴			
BMI	−0.457	Glutamic acid	−0.188
Lifetime alcohol	−0.671	Hexoses	−0.239
Diet score	0.429	SM C16:1	0.203
Physical activity	−0.093	SM(OH) C14:1	0.187
Smoking	−0.047	SM(OH) C22:2	0.219
Hepatitis infection	0.019	LysoPC aC28:1	0.186
Diabetes	−0.382	PC acC30:2	0.184

¹LysoPC, lyso-phosphatidylcholine; PC aa, diacyl-phosphatidylcholine; PC ac, acyl-alkylphosphatidylcholine; PLS, partial least squares; SM, sphingomyelin.

²Metabolite variables contributing to each PLS factor were selected based on extreme loading values, i.e., below or above the 2.5th and 97.5th percentiles.

³Results from the main analysis using residuals based on country (X- and M-sets) and batch (M-set only) ($n = 294$, X-set = 7, M-set = 133).

⁴Results from the sensitivity analysis using the same residuals as in the main analysis but excluding the liver function score from the M-set ($n = 294$, X-set = 7, M-set = 132).

regression (34). This technique extracts linear combinations, referred to as PLS factors, of predictors (the X-set of lifestyle variables) and responses (the M-set of metabolites), allowing a simultaneous decomposition of both sets with the aim of maximizing their covariance (34). The mathematical and computational details of the PLS method and its applicability have been described thoroughly in our previous study on HCC within EPIC (35). The number of PLS factors to retain was selected after carrying out a 7-fold cross-validation to minimize the predicted residual sum of squares statistic, a measure of PLS performance. Details of the process can be found elsewhere (35). PLS factor loadings, i.e., the coefficients quantifying how much each original variable contributes to the PLS factor, characterized each extracted lifestyle and metabolic profile. As the M-set was particularly dense in metabolite variables, the interpretation of the metabolic profile mainly focused on those most significantly contributing to the PLS component, reporting variables with loading values lower than the 2.5th and larger than the 97.5th percentile. A sensitivity PLS analysis was performed by excluding the liver function score from the M-set. This was done to investigate the performance of the PLS-obtained signatures, ruling out any potential bias that may arise when including a composite score based on enzyme biomarker levels that was closely associated with HCC (32). Additionally, analyses were performed excluding pairs of cases and controls where cases were diagnosed within the first 2 and 4 y following blood collection and excluding case sets where cases and/or controls were hepatitis positive (**Supplemental Table 2**).

Conditional logistic regression analyses

The associations between the modified HLI as well as the PLS scores reflecting the lifestyle and metabolic signatures and the

HCC risk were evaluated separately in conditional logistic regression models. ORs and their 95% CIs were computed to express a change in HCC risk reflecting a 1-SD increase in the index and the PLS scores, respectively. The models with the modified HLI and the main PLS analysis were unadjusted, whereas the models from the sensitivity PLS analysis were adjusted for the liver function score. These conditional logistic regression models were used to build the receiver operating characteristic (ROC) curves and determine their AUCs, to evaluate the predictive performance of the models. The DeLong test was used to evaluate the difference between AUCs where applicable. Associated statistics for sensitivity, specificity, and accuracy were computed for a cut-off point, selected as the minimal distance between the ROC curve and the upper left corner of the diagram. To account for the nested case-control design, the positive and negative predictive values (36) were computed by including the HCC prevalence within EPIC ($\pi = 0.0004$) calculated over an 8-y period corresponding to the mean follow-up time (from 1992–2002 to 2010) from a total of 477,206 participants included for case identification after relevant exclusions where 191 HCC cases were identified (29).

The different analytical steps described herein are outlined in **Figure 1**. All statistical tests were two-sided and P values <0.05 were considered statistically significant. Statistical analyses were performed with the use of PROC PLS in SAS (Base SAS 9.4, SAS Institute Inc.) for PLS analyses and R Software version 3.3.1 (R Foundation for Statistical Computing) for PC-PR2 analysis, conditional logistic regressions, and ROC-related statistics (packages survival and OptimalCutpoints).

RESULTS

Study population characteristics by case-control status are presented in **Table 1**. Following the application of PC-PR2, totals of

TABLE 3

Multivariable ORs and 95% CIs for the association of HCC with the modified HLI and the PLS lifestyle and metabolic signatures (X- and M-scores)¹

Exposure	OR (95% CI) ²	P-value
HLI associations		
Modified HLI	0.82 (0.76, 0.89)	2.3×10^{-6}
Main PLS analysis ³		
Lifestyle signature	0.53 (0.38, 0.74)	2.6×10^{-5}
Metabolic signature	0.28 (0.18, 0.43)	8.0×10^{-9}
Sensitivity PLS analysis ⁴		
Lifestyle signature ⁵	0.69 (0.48, 1.01)	5.8×10^{-2}
Metabolic signature ⁵	0.29 (0.16, 0.52)	3.4×10^{-5}

¹HCC, hepatocellular carcinoma; HLI, healthy lifestyle index; PLS, partial least squares.

²ORs are reported for a 1-SD increase either in the modified HLI or in the PLS X- and M-scores (lifestyle and metabolic signatures). The ORs from the HLI and the main PLS analysis were unadjusted, whereas the conditional regression models in the sensitivity analysis were adjusted for the liver function score. The OR results for the lifestyle and metabolic signatures were computed from two separated models, and were not mutually adjusted, in both the main and sensitivity analyses. Cases and controls were matched on age at blood collection (± 1 y), sex, study center, date (± 2 mo) and time of day at blood collection (± 3 h), and fasting status at blood collection (< 3 , $3-6$, > 6 h); women were additionally matched on menopausal status (pre-, peri-, postmenopausal) and hormone replacement therapy.

³Results from the main analysis using residuals based on country (X- and M-sets) and batch (M-set only) ($n = 294$, X-set = 7, M-set = 133).

⁴Results from the sensitivity analysis using the same residuals as in the main PLS analysis but excluding the liver function score from the M-set ($n = 294$, X-set = 7, M-set = 132).

⁵This model was adjusted for the liver function score.

6 and 21 principal components were retained, explaining around 80% of the total variability among the modified HLI original variables and the metabolites set, respectively. Results reported in **Supplemental Figure 2** showed that the ensemble of explanatory variables accounted for 10.7% and 29.5% of the total variance within the X- and M-sets, respectively. “Country of origin” was the highest contributor to the variance, with consistently 6.2% and 13.1% in the X- and M-sets, followed by “Batch”, with 7.1% in the M-set (Supplemental Figure 2). For both the X- and M-sets, residuals on country and batch (M-set only) were computed in univariate linear regression models and used in the PLS analyses.

One PLS factor was retained after 7-fold cross-validation for PLS analysis. The score plot of this factor is depicted in **Supplemental Figure 3**. The lifestyle PLS factor identified a “healthy” behavior profile, with positive loadings for physical activity and negative loadings for BMI, lifetime alcohol consumption, and smoking (**Table 2**). The corresponding metabolic PLS factor was characterized by glutamic acid, hexoses and sphingomyelins (SMs), including SM(OH) C14:1, SM(OH) C16:1, and SM(OH) C22:2. The PLS lifestyle factor was inversely associated with HCC risk, with OR = 0.53 (95% CI: 0.38, 0.74, $P = 2.64 \times 10^{-5}$) (**Table 3**); the direction of the association was stronger than that of the modified HLI score, with OR = 0.82 (95% CI: 0.76, 0.89, $P = 2.3 \times 10^{-6}$). However, the PLS metabolic signature showed a much stronger inverse association with HCC risk, with OR = 0.28 (95% CI: 0.18, 0.43, $P = 8.0 \times 10^{-9}$) (**Table 3**). The lifestyle and metabolic signatures remained virtually unchanged after excluding case sets where cases were diagnosed within the 2 and 4 y after blood collection. The association with HCC risk remained

strong, with OR = 0.46 (95% CI: 0.32, 0.70, $P = 3.03 \times 10^{-5}$) for lifestyle and OR = 0.27 (95% CI: 0.17, 0.43, $P = 6.03 \times 10^{-8}$) for metabolic signature, respectively, for the lag time of 2 y, and OR = 0.54 (95% CI: 0.37, 0.80, $P = 3.64 \times 10^{-4}$) and OR = 0.30 (95% CI: 0.19, 0.50, $P = 1.69 \times 10^{-6}$) for the lag time of 4 y. Similarly, after excluding hepatitis-positive case sets, the extracted metabolic signature remained inversely related to HCC risk, with OR = 0.33 (95% CI: 0.21, 0.52, $P = 2.50 \times 10^{-6}$) (Supplemental Table 2).

Results from the sensitivity analysis (**Table 3**) that was performed excluding the liver function score from the M-set yielded similar results, with the metabolic profile positively associated with sphingomyelins [SM C16:1, SM(OH) C14:1, and SM(OH) C22:2] and phosphatidylcholines [lyso-phosphatidylcholine (LysoPC) aC28:1 and acyl-alkylphosphatidylcholine (PC ae) C 30:2] and negatively associated with glutamic acid and hexoses. The lifestyle signature showed positive loadings for the diet score and negative loadings for BMI, lifetime alcohol intake, and diabetes. The PLS factor was associated with a decrease in HCC risk through both its lifestyle and metabolic profiles, with OR = 0.69 (95% CI: 0.48, 1.01, $P = 5.79 \times 10^{-2}$) and OR = 0.29 (95% CI: 0.16, 0.52, $P = 3.36 \times 10^{-5}$), respectively.

ROC parameters, including AUC values, sensitivity, specificity, accuracy, positive predictive value, and negative predictive value, are reported in **Table 4** for the modified HLI and for PLS lifestyle and metabolic signatures both in main and sensitivity analyses. The AUC values for the metabolic signatures were consistently higher than their lifestyle counterparts, with AUC = 0.74 (95% CI: 0.69, 0.80) and AUC = 0.64 (95% CI: 0.57, 0.70), respectively, in the main analysis and with AUC = 0.83 (95% CI: 0.79, 0.88) and AUC = 0.78 (95% CI: 0.73, 0.84) in the sensitivity analysis. The AUC for the liver function was 0.77 (95% CI: 0.72, 0.82), and including the metabolic signature from the sensitivity analysis resulted in a statistically significant increase of the HCC predictability to AUC = 0.83 (95% CI: 0.79, 0.88) with $P_{\text{DeLong}} = 0.003$. Additionally, the metabolic signature had a higher AUC value than the modified HLI [AUC = 0.67 (95% CI: 0.61, 0.73)]. Model performance of the metabolic signature was minimally affected when hepatitis-positive pairs were excluded (AUC = 0.73 (95% CI: 0.66, 0.80)) (Supplemental Table 2).

DISCUSSION

Statement of principal findings

In this analysis, we employed a multivariate dimension reduction method to identify a metabolic signature reflecting a “healthy lifestyle” profile. The metabolic signature was associated with a significant 72% (95% CI: 57%, 82%) reduction in HCC risk, which was statistically significantly different from the 51% (95% CI: 32%, 65%) risk reduction associated with a lifestyle profile based on questionnaire data. The metabolic signature had a higher predictability with respect to HCC, with AUC = 0.74 (95% CI: 0.69, 0.80) compared with AUC = 0.64 (95% CI: 0.57, 0.70) for the lifestyle profile. These findings highlight the potential for metabolic profiling to capture data on pathophysiological processes that are associated with lifestyle exposures and significantly improve identification of at-risk individuals in the general population.

TABLE 4

AUC, sensitivity, specificity, accuracy, PPV, and NPV of ROC models (with 95% CI) from associations of the modified HLI and the PLS lifestyle and metabolic signatures (X- and M-scores) with HCC¹

	AUC	<i>P</i> value ²	Sensitivity	Specificity	Accuracy	PPV	NPV
HLI associations							
Modified HLI	0.67 (0.61, 0.73)	—	0.63 (0.54, 0.70)	0.59 (0.50, 0.67)	0.61	0.0006	0.37
Main PLS analysis ³							
Lifestyle signature	0.64 (0.57, 0.70)	—	0.54 (0.46, 0.63)	0.71 (0.63, 0.78)	0.63	0.0007	0.41
Metabolic signature	0.74 (0.69, 0.80)	—	0.61 (0.52, 0.68)	0.78 (0.70, 0.84)	0.69	0.0011	0.44
Sensitivity PLS analysis ⁴							
Liver function score (LFS)	0.77 (0.72, 0.82)	—	0.70 (0.61, 0.77)	0.75 (0.67, 0.82)	0.72	0.0011	0.43
LFS + lifestyle signature	0.78 (0.73, 0.84)	0.407	0.73 (0.65, 0.80)	0.73 (0.65, 0.80)	0.73	0.0011	0.42
LFS + metabolic signature	0.83 (0.79, 0.88)	0.003	0.74 (0.66, 0.81)	0.80 (0.73, 0.86)	0.77	0.0015	0.45

¹The estimates for the lifestyle and metabolic signatures were computed from two separated models, and were not mutually adjusted, both in main and sensitivity analyses. HCC, hepatocellular carcinoma; HLI, healthy lifestyle index; LFS, liver function score; NPV, negative predicted value; PPV, positive predicted value; PLS, partial least squares; ROC, receiver operating characteristic curve.

²*P* value from the DeLong test comparing the AUCs of the models including only the LFS with the models also including the lifestyle and the metabolic signature, respectively.

³Results from the main analysis using residuals based on country (X- and M-sets) and batch (M-set only) (*n* = 294, X-set = 7, M-set = 133).

⁴Results from the sensitivity analysis excluding the liver function score from the M-set (*n* = 294, X-set = 7, M-set = 132).

Strengths and limitations of the study

To our knowledge, this study is the first to derive metabolic signatures of healthy lifestyle behaviors and at the same time relate the signatures to the risk of HCC risk, thus offering objective information on the mechanistic processes involved. Our findings showed that metabolic profiles substantially improved discrimination of at-risk individuals compared with questionnaire-derived data and liver function tests (14). Hitherto, large-scale prospective studies examining the association between combined lifestyle factors or patterns of healthy behaviors with chronic disease (37), including cancer subtypes (24, 38), have relied on indices derived from self-reported questionnaire-based variables (11), without including metabolites associated with healthy habits. The methodology described herein was tailored to HCC by examining a restricted set of 7 exposures from the modified HLI that have previously been associated with HCC risk (4–7, 39, 40) with the aim of identifying robust metabolic signals.

The derived metabolic signature remained strongly associated with a 71% decrease in HCC risk after separating out the liver function score from the M-set. In addition, the metabolic signature increased the prediction of HCC risk when it was added on top of the liver function score, with a higher AUC of 0.83 (95% CI: 0.79, 0.88), compared with AUC of 0.77 (95% CI: 0.72, 0.82), suggesting that metabolites can add further information to models that relate liver function enzymes to HCC risk only. Additional data from untargeted metabolomics might improve this prediction.

A limitation of the study is its small sample size, since HCC is a rare malignancy and few prospective cohorts have a sufficient number of incident cases with prediagnostic blood samples. As a consequence, both cases and controls were used in the PLS analysis to extract lifestyle and metabolic signatures, though the analysis was carried out blindly with regard to the case-control status and blood samples were obtained prior to cancer diagnosis. Despite the prospective nature of our design, we cannot rule out potential reverse causation, as the concentrations of some metabolites may have been modified by an underlying subclinical carcinogenic process. Yet, after performing sensitivity analyses

excluding sets of cases and controls where cases were diagnosed within the first 2 (and 4) y after recruitment, the metabolic signatures remained stable and strongly inversely associated with HCC risk. It is noteworthy that we observed lower hepatitis positivity (~10–15%) in our European population relative to the general population, where rates are as high as 30% for hepatitis B and/or C (41). In our analysis, missing values for hepatitis infection (~25%) were imputed by an EM algorithm (30). Despite the small sample size, a sensitivity analysis was conducted excluding hepatitis-negative cases and controls, and the results indicated that although hepatitis infection plays a prominent role in HCC occurrence, signatures related to lifestyle habits can explain HCC carcinogenesis. Lastly, this study was conducted in a nested case-control sample within a European population. The associations between HLI and HCC risk was marginally heterogeneous by sex (results not shown) but, given the limited sample size, this could not be investigated further.

Comparison with other studies and potential mechanisms

A number of studies have investigated the association between individual lifestyle (13, 17, 20, 23), anthropometric (42) and food items (20, 43), with concentrations of blood metabolites relying on high-resolution nuclear magnetic resonance or mass spectroscopy, yet only a few have studied metabolic signatures of a number of exposures simultaneously. More recently, metabolite phenotyping has been used to objectively assess dietary patterns by identifying metabolite profiles of healthy eating patterns (20, 44).

In our study, the metabolites contributing the greatest loadings to the HLI-related metabolic signatures were glutamic acid, hexoses, sphingomyelins [SM C16:1, SM(OH) C14:1, C16:1, and C22:2], and phosphatidylcholines [LysoPC aC28:1, diacyl-phosphatidylcholine (PC aa) C32:1, and PC aeC30:2]. These patterns remained virtually unchanged in the sensitivity analysis where the score of liver enzymes was removed from the set. Results from two studies conducted in a component subcohort of EPIC (19, 42) showed that PC aes and

diacyl-phosphatidylcholines (PC aas) were, respectively, negatively and positively associated with BMI and other anthropometric markers of obesity, including waist and hip circumferences. PC aas and LysoPCs were inversely associated with dietary fiber sources, although LysoPCs were positively related to physical activity (19).

Ethanol has been hypothesized to disrupt the metabolism of phospholipids by inducing lipogenesis in the liver tissues (45), leading to hepatic injuries and precancerous lesions. High intakes of alcohol can result in the activation of the enzyme acid sphingomyelinase, which likely accelerates the catabolism of SMs into ceramide and PCs (46), and in turn leads to hepatotoxicity and HCC (47). Glutamic acid concentrations were positively associated with alcohol intake in Japanese men of the Tsuruoka Metabolomics Cohort Study (13).

SM(OH) C22:2 was negatively associated with smoking in women in KORA S4 (23) and with alcohol intake in the CARLA study (22). PC aaC32:1 was positively related to alcohol use in CARLA (22), type 2 diabetes (T2D) in EPIC-Potsdam (17), and high prevalence of smoking in men (22) and low prevalence of smoking in women (23). Free radicals contained in cigarette smoke promote lipid peroxidation (48), which can result in liver disease (49). Smoking also lowers levels of PC aas and raises those of PC aas, impeding lipid remodeling in membranes, which results in inflammation (50), and may in turn lead to liver injury (51).

A number of studies based on targeted (17) and untargeted (52) metabolomic data identified hexoses as biomarkers of diabetes and related it to increased risk of T2D. In our study, the sum of hexoses was negatively associated with the metabolic signature. Hexoses relate to hyperglycemia, which can be monitored in diabetic patients, who are at higher risk of developing HCC (4). In a study using metabolomics for biomarker discovery of T2D (17), SM C16:1 was associated with decreased T2D risk and was previously identified as a prediabetes biomarker (53).

Lastly, some of the serum liver enzymes in the liver function score have previously been related to high ethanol intake (13), hepatitis infection (18), and low physical exercise (7), and can be used as a panel to detect abnormal conditions, such as hepatitis and cirrhosis, preceding HCC (51). A panel of metabolites, including the hexoses, the identified phospholipids, and glutamic acid, can be measured in at-risk patients to detect their propensity to develop HCC, in addition to key liver enzymes.

Conclusions

Our findings indicate that a specific panel of metabolites linked to healthy lifestyle habits was strongly associated with HCC and can substantially improve HCC risk prediction beyond questionnaire-derived variables and liver function tests. The identified metabolites may also offer insights into the underlying biological mechanisms driving HCC development. Replication of these findings in an independent setting is now warranted to enhance the understanding of the relation between lifestyle exposures and health outcomes through metabolic profiling.

The authors thank Bertrand Hémon and Carine Biessy from IARC for their help with issues related to data management.

The authors' responsibilities were as follows—NA, PF, and VV: conceptualized the study and defined the analytical strategy; NA: was responsible for the statistical analyses, provided preliminary interpretation of the findings and developed the first draft of the manuscript; NA, MJG, VV, and PF:

wrote the paper; PF, VV, DCT, ML, MS, MJ, A Scalbert, and MJG contributed to the drafting of the manuscript; VC, TP, CB, M-CBR, TMS, AM, HB, A Sundkvist, TK, RT, KO, ER, and MJG substantially contributed to the interpretation of results and critically revised the content of the manuscript; and all authors: critically helped in the interpretation of results, provided relevant intellectual input, and read, revised, and approved the final manuscript. NA was financially supported by the Université Claude Bernard Lyon I through a doctoral fellowship awarded by the EDISS (Ecole Doctorale Interdisciplinaire Sciences Santé) doctoral school to complete her PhD work. NA also holds a grant from the Fondation de France (FdF) supporting her postdoctoral research (grant number: 00069254). The data on the EPIC-Hepatobiliary dataset was generated through support from the French National Cancer Institute (L'Institut National du Cancer; INCA) (grant number 2009–139; PI: MJ). The work undertaken by DCT reported in this publication was supported by the National Institutes of Health under award number P01 CA196559. RT is a co-principal investigator of the EPIC-Oxford cohort whose work is supported by Cancer Research UK under grant number C8221/A19170. None of the authors reported a conflict of interest related to the study.

REFERENCES

1. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* 2013;132:1133–45.
2. Mittal S, El-Serag HB. Epidemiology of HCC: consider the population. *J Clin Gastroenterol* 2013;47:S2–6.
3. Bertuccio P, Turati F, Carioli G, Rodriguez T, La Vecchia C, Malvezzi M, Negri E. Global trends and predictions in hepatocellular carcinoma mortality. *J Hepatol* 2017;67:302–9.
4. Goma A-I. Hepatocellular carcinoma: epidemiology, risk factors and pathogenesis. *World J Gastroenterol* 2008;14:4300–8.
5. Chiang C-H, Lu C-W, Han H-C, Hung S-H, Lee Y-H, Yang K-C, Huang K-C. The relationship of diabetes and smoking status to hepatocellular carcinoma mortality. *Medicine (Baltimore)* 2016;95:e2699.
6. Turati F, Trichopoulos D, Polesel J, Bravi F, Rossi M, Talamini R, Franceschi S, Montella M, Trichopoulou A, La Vecchia C et al. Mediterranean diet and hepatocellular carcinoma. *J Hepatol* 2014;60:606–11.
7. Berzigotti A, Saran U, Dufour J-F. Physical activity and liver diseases. *Hepatology* 2016;63:1026–40.
8. Gupta A, Das A, Majumder K, Arora N, Mayo HG, Singh PP, Beg MS, Singh S. Obesity is independently associated with increased risk of hepatocellular cancer-related mortality. *Am J Clin Oncol [Internet]* 2017;1. Available from: <http://insights.ovid.com/crossref?an=00000421-900000000-98965>.
9. Tavoschi L, Duffell E, De H, Gomes C, Amato-Gauci A, Hofstraat S, Falla A, Veldhuijzen I, Hahné S, Van Benthem B. Systematic review on hepatitis B and C prevalence in the EU/EEA [Internet]. [cited 2018 Feb 17]; Available from: <https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/systematic-review-hepatitis-B-C-prevalence.pdf>.
10. Center for Disease Control—CDC. Viral Hepatitis Surveillance — United States, 2015 [Internet]. Available from: <https://www.cdc.gov/hepatitis/statistics/2015surveillance/pdfs/2015hepsurveillancerept.pdf>.
11. Day N, Ferrari P. Some methodological issues in nutritional epidemiology. *IARC Sci Publ* 2002;156:5–10.
12. Cano A, Alonso C. Deciphering non-alcoholic fatty liver disease through metabolomics. *Biochem Soc Trans* 2014;42:1447–52.
13. Harada S, Takebayashi T, Kurihara A, Akiyama M, Suzuki A, Hatakeyama Y, Sugiyama D, Kuwabara K, Takeuchi A, Okamura T, et al. Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. *Environ Health Prev Med* 2016;21:18–26.
14. Nicholson JK, Holmes E, Elliott P. The metabolome-wide association study: a new look at human disease risk factors. *J Proteome Res* 2008;7:3637–8.
15. Stepien M, Chajes V, Romieu I. The role of diet in cancer: the epidemiologic link. *Salud Publica Mex* 2016;58:261–73.
16. Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, Olsen A, Tjønneland A, Dragsted LO. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics* 2015;11:1376–80.

17. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG, Fritsche A, Häring HU, De Angelis MH, Peters A, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 2013;62:639–48.
18. Saito T, Sugimoto M, Okumoto K, Haga H, Katsumi T, Mizuno K. Serum metabolome profiles characterized by patients with hepatocellular carcinoma associated with hepatitis B and C. *World J Gastroenterol* 2016;22:6224–34.
19. Floegel A, Wientzek A, Bachlechner U, Jacobs S, Drogan D, Prehn C, Adamski J, Krumsiek J, Schulze MB, Pischon T, et al. Linking diet, physical activity, cardiorespiratory fitness and obesity to serum metabolite networks: findings from a population-based study. *Int J Obes (Lond)* 2014;38:1388–96.
20. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, Albanes D, Gu F, Kontto J, Lassale C, et al. Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr* 2017;105:450–65.
21. Carayol M, Leitzmann M, Ferrari P, Zamora-Ros R, Achaintre D, Schmidt JA, Travis RC, Jenab M, Key TJ, Scalbert A, et al. Blood metabolic signatures of body mass index: a targeted metabolomics study in the EPIC cohort. *J Proteome Res* 2017;16:3137–46.
22. Lacruz ME, Kluttig A, Tiller D, Medenwald D, Giegling I, Rujescu D, Prehn C, Adamski J, Frantz S, Greiser KH, et al. Cardiovascular risk factors associated with blood metabolite concentrations and their alterations during a 4-year period in a population-based cohort. *Circ Cardiovasc Genet* 2016;9:487–94.
23. Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, Perstorfer K, Jaremek M, Roemisch-Margl W, Rathmann W, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med* 2013;11:60.
24. McKenzie F, Biessy C, Ferrari P, Freisling H, Rinaldi S, Chajès V, Dahm CC, Overvad K, Dossus L, Lagiou P, et al. Healthy lifestyle and risk of cancer in the European Prospective Investigation into Cancer and Nutrition cohort study. *Medicine (Baltimore)* 2016;95:e2850.
25. McKenzie F, Ferrari P, Freisling H, Chajès V, Rinaldi S, de Batlle J, Dahm CC, Overvad K, Baglietto L, Dartois L, et al. Healthy lifestyle and risk of breast cancer among postmenopausal women in the European Prospective Investigation into Cancer and Nutrition cohort study. *Int J Cancer* 2015;136:2640–8.
26. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondière UR, Hémon B, Casagrande C, Vignat J, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
27. Slimani N, Deharveng G, Unwin I, Southgate DAT, Vignat J, Skeie G, Salvini S, Parpinel M, Møller A, Ireland J, et al. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr* 2007;61:1037–56.
28. Spencer EA, Roddam AW, Key TJ. Accuracy of self-reported waist and hip measurements in 4492 EPIC-Oxford participants. *Public Health Nutr* 2004;7:723–7.
29. Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, Achaintre D, Assi N, Tjønneland A, Overvad K, et al. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: findings from a prospective cohort study. *Int J Cancer* 2016;138:348–60.
30. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 1977;39:1–38.
31. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, Moret NCO, Scalbert A, Rinaldi S, Ferrari P. Reliability of serum metabolites over a two-year period: a targeted metabolomic approach in fasting and non-fasting samples from EPIC. *PLoS One* 2015;10:1–10.
32. Fedirko V, Trichopolou A, Bamia C, Duarte-Salles T, Trepo E, Aleksandrova K, Nöthlings U, Lukanova A, Lagiou P, Boffetta P, et al. Consumption of fish and meats and risk of hepatocellular carcinoma: the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol* 2013;24:2166–73.
33. Fages A, Ferrari P, Monni S, Dossus L, Floegel A, Mode N, Johansson M, Travis RC, Bamia C, Sánchez-Pérez MJ, et al. Investigating sources of variability in metabolomic data in the EPIC study: the Principal Component Partial R-square (PC-PR2) method. *Metabolomics* 2014;10:1074–83.
34. Tenenhaus M. *La régression PLS*. Paris: Technip; 1998.
35. Assi N, Fages A, Vineis P, Chadeau-Hyam M, Stepien M, Duarte-Salles T, Byrnes G, Boumaza H, Knüppel S, Kühn T, et al. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis* 2015;30:743–53.
36. van Zaane B, Vergouwe Y, Donders AR, Moons KGM. Comparison of approaches to estimate confidence intervals of post-test probabilities of diagnostic test results in a nested case-control study. *BMC Med Res Methodol* 2012;12:166.
37. Kurth T, Moore SC, Gaziano JM, Kase CS, Stampfer MJ, Berger K, Buring JE. Healthy lifestyle and the risk of stroke in women. *Arch Intern Med* 2006;166:1403.
38. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjønneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. *BMJ* 2010;341:c5504.
39. Yang W-S, Va P, Bray F, Gao S, Gao J, Li H-L, Xiang Y-B. The role of pre-existing diabetes mellitus on hepatocellular carcinoma occurrence and prognosis: a meta-analysis of prospective cohort studies. *PLoS One* 2011;6:e27326.
40. Testino G, Leone S, Borro P. Alcohol and hepatocellular carcinoma: a review and a point of view. *World J Gastroenterol* 2014;20:15943–54.
41. Duddempudi AT, Bernstein DE. Hepatitis B and C. *Clin Geriatr Med* 2014;30:149–67.
42. Bachlechner U, Floegel A, Steffen A, Prehn C, Adamski J, Pischon T, Boeing H. Associations of anthropometric markers with serum metabolites using a targeted metabolomics approach: results of the EPIC-Potsdam study. *Nutr Diabetes* 2016;6:e215.
43. Floegel A, von Ruesten A, Drogan D, Schulze MB, Prehn C, Adamski J, Pischon T, Boeing H. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr* 2013;67:1100–8.
44. Garcia-Perez I, Gibson Bsc R, Chambers ES, Frost G, Posma JM, Nicholson JK, Holmes E, Elliott P, Hansen TH, Vestergaard H, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol* 2017;5:184–95.
45. You M, Fischer M, Deeg MA, Crabb DW. Ethanol induces fatty acid synthesis pathways by activation of sterol regulatory element-binding protein (SREBP). *J Biol Chem* 2002;277:29342–7.
46. Reichel M, Hönig S, Liebisch G, Lüth A, Kleuser B, Gulbins E, Schmitz G, Kornhuber J. Alterations of plasma glycerophospholipid and sphingolipid species in male alcohol-dependent patients. *Biochim Biophys Acta* 2015;1851:1501–10.
47. Deaciuc IV, Nikolova-Karakashian M, Fortunato F, Lee EY, Hill DB, McClain CJ. Apoptosis and dysregulated ceramide metabolism in a murine model of alcohol-enhanced lipopolysaccharide hepatotoxicity. *Alcohol Clin Exp Res* 2000;24:1557–65.
48. Girotti AW. Mechanisms of lipid peroxidation. *J Free Radic Biol Med* 1985;1:87–95.
49. Li S, Tan H-Y, Wang N, Zhang Z-J, Lao L, Wong C-W, Feng Y. The role of oxidative stress and antioxidants in liver diseases. *Int J Mol Sci* 2015;16:26087–124.
50. González-Reimers E, Santolaria-Fernández F, Martín-González MC, Fernández-Rodríguez CM, Quintero-Platt G. Alcoholism: a systemic proinflammatory condition. *World J Gastroenterol* 2014;20:14660–71.
51. Nagata K, Suzuki H, Sakaguchi S. Common pathogenic mechanism in development progression of liver injury caused by non-alcoholic or alcoholic steatohepatitis. *J Toxicol Sci* 2007;32:453–68.
52. Drogan D, Dunn WB, Lin W, Buijsse B, Schulze MB, Langenberg C, Brown M, Floegel A, Dietrich S, Rolandsson O, et al. Untargeted metabolic profiling identifies altered serum metabolites of type 2 diabetes mellitus in a prospective, nested case control study. *Clin Chem* 2015;61:487–97.
53. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012;8:615.